RNA-RNA INTERACTION PREDICTION:
PARTITION FUNCTION AND BASE PAIR PAIRING PROBABILITIES

FENIX W.D. HUANG[1], JING QIN[1], CHRISTIAN M. REIDYS[1,⋆] AND PETER F. STADLER[2]

ABSTRACT. In this paper, we study the interaction of an antisense RNA and its target mRNA, based on the model introduced by Alkan *et al.* (Alkan *et al.*, J. Comput. Biol., Vol:267–282, 2006). Our main results are the derivation of the partition function [11] (Chitsaz *et al.*, Bioinformatics, to appear, 2009), based on the concept of tight-structure and the computation of the base pairing probabilities. This paper contains the folding algorithm rip which computes the partition function as well as the base pairing probabilities in $O(N^4 M^2) + O(N^2 M^4)$ time and $O(N^2 M^2)$ space, where N, M denote the lengths of the interacting sequences.

RNA-RNA interaction, joint structure, dynamic programming, partition function, base pairing probability, loop, RNA secondary structure.

1. INTRODUCTION

The discovery of small RNAs that bind to their target mRNAs in order to prohibit their translation and down-regulate the expression levels of corresponding genes has drawn a lot of attention in the RNA world [21]. Studies have shown that many RNA-RNA interactions play a significant role in different cellular processes, such as mediate pseudouridylation and methylation of rRNA [4], nucleotide insertion into mRNAs [6], splicing of pre-mRNA [35] and translation control or plasmid replication control [5, 12, 18].

Regulatory RNAs constitute a subclass of the antisense RNA family; encompassing the snRNAs, gRNAs and snoRNAs that play a role in the context of rRNA modification, RNA editing, mRNA spicing and plasmid copy-number regulation. In addition, antisense RNAs are synthesized for

studying specific gene functions. Since the first published result on natural antisense RNAs which regulate gene expression in C. elegans [25, 34, 13, 27], Drosophila [24], and other organisms [31], the problem of predicting how two nucleic acid strands interact–the so called RNA-RNA interaction problem (RIP)–has come into focus.

As observed by Alkan *et al.* [2], the RIP is NP-complete. The actual argument constitutes an extension of the work of Akutsu [1] derived in the context of single RNA secondary structure prediction problems with pseudoknots. As in Rivas and Eddys pseudoknot folding algorithm [29] the general idea here is to consider specific classes of interactions, that can be computed via dynamic programming routines. There are several other methods that consider somewhat restricted versions of the RNA-RNA interaction. For instance, one method concatenates the two interacting sequences and subsequently employs a slightly modified standard secondary structure folding algorithm. The algorithms RNAcofold [14, 7], pairfold [3] and NUPACK [28] subscribe to this strategy. However, this approach cannot predict important motifs in RIPs, as for instance kissing hairpin loops. The concatenation idea has also been employed using the pseudoknot folding algorithm of Rivas and Eddy [29]. The resulting algorithm, however, does still not generate all relevant interaction structures [11, 26]. An alternative line of thought is to neglect all internal base-pairings in either strand and to compute the minimum free energy (mfe) secondary structure for their hybridization under this constraint. For instance, RNAduplex follows this line of thought making it formally equivalent to the classic secondary structure folding algorithm of Waterman [32, 15, 33, 30]. Furthermore we have the algorithm RNAup [23, 22] which uses the Alkan's model, allowing for *one* interaction region having unbranched interactions within any loop. RNAup can therefore capture single but not multiple kissing hairpins. Finally there is IntaRNA [8] facilitating the efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions.

Alkan *et al.* [2] derived a mfe algorithm for predicting the joint secondary structure of two interacting RNA molecules with polynomial time complexity. Here "joint structure", see Fig. 1 for example, means that the intramolecular structures of each molecule are pseudoknot-free, the intermolecular binding pairs are noncrossing and there exist no so called "zig-zags" (see Section 1 for details). Zig-zags are sometimes referred to as tangles.

Recently, Chitsaz et.al. [11] presented a dynamic programming algorithm which computes the partition function in $O(N^6)$ time. The key point for passing from the mfe folding of Alkan [2] to the partition function is a unique grammar by which each interaction structure can be generated.
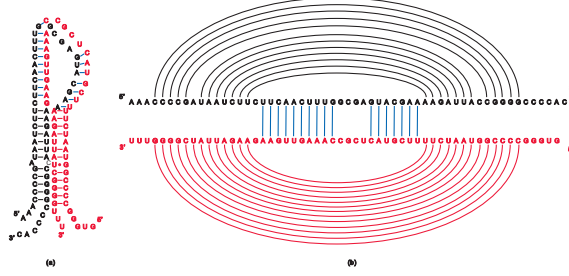
FIGURE 1. Natural joint structure between small RNA molecules CopA(antisense) and CopT(target) in E.coli [2].

The dynamic programming routine for the partition function of RNA secondary structures is due to McCaskill [20] and can be outlined as follows: the free energy of a secondary structure is assumed additive in terms of its loops $F(S) = \sum_{L \in S} F_L$, where F_L denotes the free energy of a loop,L. The additivity of the free energy translates itself into the multiplicativity in the contributions to the partition function Q defined by $Q = \sum_S e^{-F(S)/kT}$, where Q is the sum over all the secondary structures S of length M. This factorization of terms can be realized by introducing $Q^b(i,j)$, where the sum is taken over all substructures $S[i,j]$ on the segment $[i,j]$ for which $(S[i], S[j]) \in S[i,j]$ and $Q^s(i,j)$ for all the configurations on $[i,j]$, irrespective of whether or not i,j are connected. In particular, we have $Q^s(1,M) = Q$. Consequently, we arrive at the recursion, see Fig. 3

$$(1.1) \qquad Q^s(i,j) = 1 + \sum_{h,\ell} Q^s(i, h-1) Q^b(h, \ell).$$

Let us next recall the basic loops-types upon which the partition function and energy parameters [19] of RNA secondary structures are based:

(1) a *hairpin*-loop ($\mathsf{Ha}(i,j)$), is a pair $((i,j), [i+1, j-1])$, where (i,j) is an arc and $[i+1, j-1]$ is an interval, i.e. a sequence of consecutive vertices $(i, i+1, \ldots, j-1, j)$, having energy parameter $e^{-G^{\mathsf{Ha}}(i,j)/kT}$.

(2) an *interior*-loop ($\mathsf{Int}(i_1, j_1; i_2, j_2)$), is a sequence $((i_1, j_1), [i_1+1, i_2-1], (i_2, j_2), [j_2+1, j_1-1])$, where (i_2, j_2) is nested in (i_1, j_1) having the energy parameter $e^{-G^{\mathsf{Int}}(i_1, j_1; i_2, j_2)/kT}$

(3) a *multi*-loop ($\mathsf{M}(i_0, j_0)$), see Fig.2, is a sequence

$$(1.2) \qquad ([i_0, i_1-1], ((i_1, j_1), [i_1+1, j_1-1]), \ldots, ((i_t, j_t), [i_t+1, j_t-1]), [j_t+1, j_0])$$

having energy parameter $e^{-(\alpha_1 + \alpha_2(t+1) + \alpha_3 c_2)/kT}$, where $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}$, t is the number of $R[i_0+1, j_0-1]$-maximal arcs inside $R[i_0, j_0]$ and c_2 is the number of isolated vertices contained in $[i_0, j_0]$.

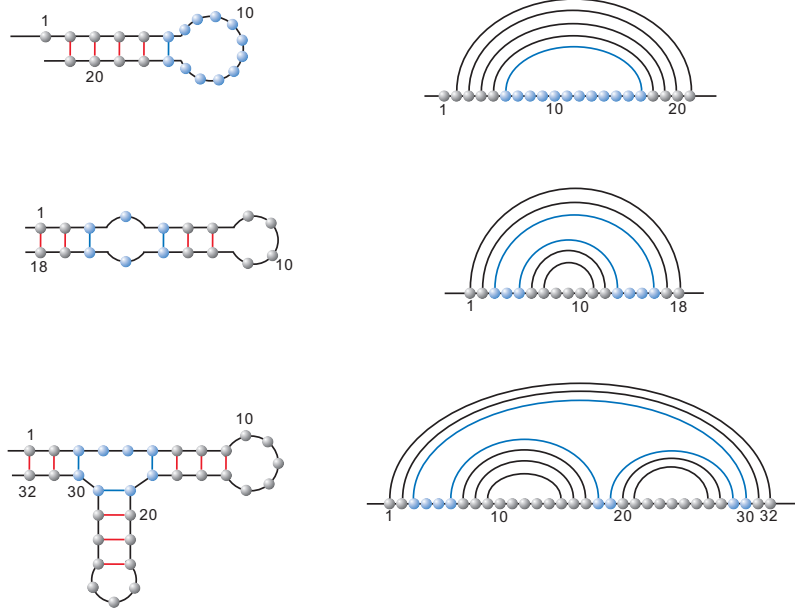Based on the above loop-energies, we obtain the following recursion for $Q^b(i,j)$

FIGURE 2. The standard loop-types for RNA secondary structures: hairpin-loop (top), interior-loop (middle) and multi-loop (bottom).

$$Q^b(i,j) = e^{-G^{\mathsf{Ha}}(i,j)/kT} + \sum_{k_1,k_2} e^{-G^{\mathsf{Int}}(i,j,k_1,k_2)/kT}$$
$$+ \sum_{\ell} Q^{m1}(i+1,\ell)Q^m(\ell+1,j-1)e^{-(\alpha_1+2\alpha_2)/kT},$$

where

$$Q^{m1}_{i,j} = \sum_{i \le \ell < j} Q^b(k,j)e^{-(\alpha_2+\alpha_3(\ell-i))/kT}$$
$$Q^m(i,j) = \sum_{i \le \ell < j} Q^{m1}(i,k)(Q^m(\ell+1,j)+e^{-\alpha_3(j-\ell)/kT}).$$

The key idea in this paper, which eventually leads to the derivation of both: the partition function as well as the base pairing probabilities, is the concept of a "tight structure", introduced in Section 2. The tight structure plays a central role in our grammar and is the main tool for obtaining the base pairing probabilities. This paper includes the folding algorithm rip, which derives the
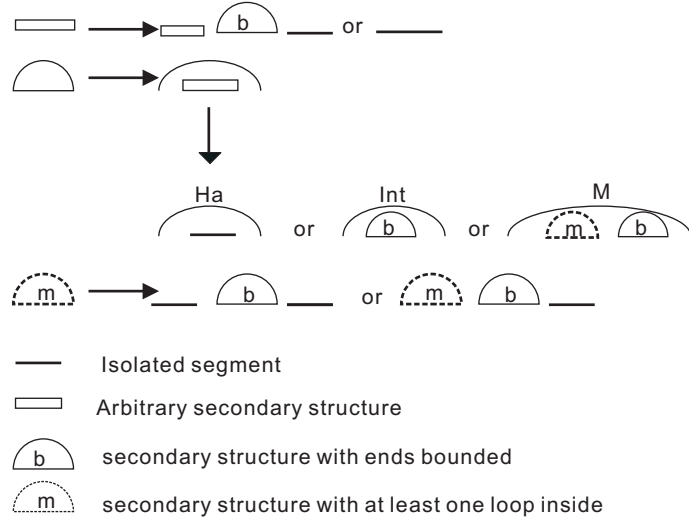
FIGURE 3. The unique decomposition of secondary structures.

partition function as well as the base pairing probabilities in $O(N^4M^2) + O(N^2M^4)$ time and $O(N^2M^2)$ space. The source code of rip is available upon request.

2. COMBINATORICS OF INTERACTION STRUCTURES

In this section we discuss some combinatorial properties of RNA interaction structures. The key idea introduced here is that of a tight structure. The main results of this section are:
• there exist only four "types" of tight structures
• given a joint structure $J(i, j; h, \ell)$, each interaction bond $(R[i_0], S[j_0]) \in J(i, j; h, \ell)$ is contained in a unique $J(i, j; h, \ell)$-tight structure
• each joint structure uniquely decomposes into a sequence of tight structures and secondary structure segments
• there exists a unique (but not canonical) decomposion of a tight structure.

Let us begin by making precise what we mean by interaction structures. Suppose we are given two diagrams [16, 17, 9, 10], R and S of length N and M, respectively. Let $R[i]$ and $S[i]$ denote

the vertex i of R and S, respectively. We shall assume that $R[1]$ denotes the $5'$ end of R and $S[1]$ denotes the $3'$ end of S as RNA sequences. The induced subgraph of S with respect to the subsequence $(S[i], \ldots, S[j])$ is denoted by $S[i,j]$. In particular, $S[i,i] = S[i]$ and $S[i, i-1] = \varnothing$. A complex $C(R, S, I)$ is a graph consisting of R, S and a set of arcs of the form $(R[i], S[j])$, I, see Fig. 4. We shall represent a complex C by drawing R on top of S with the R-arcs in the upper, the S-arcs in the lower halfplane and I-arcs vertical. Given a complex C, a subcomplex is the subgraph of C, induced by $R[i_1, j_1]$ and $S[i_2, j_2]$.
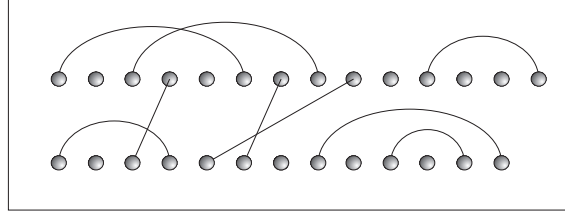

FIGURE 4. A complex C induced by $R[1, 14]$ and $S[1, 13]$.

An arc is called interior if its start and endpoint are both contained in either R or S and exterior, otherwise. Let $\prec_1$ be the partial order $\prec_1$ over the set of interior arcs, given by

$$(2.1) \qquad (S[i_1], S[j_1]) \prec_1 (S[i_2], S[j_2]) \quad \Longleftrightarrow \quad i_2 < i_1 < j_1 < j_2.$$

Similarly, let $\prec_2$ denote the partial order over the set of exterior arcs

$$(2.2) \qquad (S[i_1], R[j_1]) \prec_2 (S[i_2], R[j_2]) \quad \Longleftrightarrow \quad i_1 < i_2, \ j_1 < j_2.$$

Given an external arc, $(R[i], S[j])$, an interior arc $(R[i_1], R[j_1])$ is called its R-ancestor if $i_1 < i < j_1$ and $(S[i_2], S[j_2])$ is the S-ancestor of $(R[i], S[j])$ if $i_2 < j < j_2$, respectively. We call $(R[i], S[j])$ the descendant of $(R[i_1], R[j_1])$ and $(S[i_2], S[j_2])$ and the sets of R-ancestors and S-ancestors of $(R[i], S[j])$ are denoted by $A_R(R[i], S[j])$ and $A_S(R[i], S[j])$. The $\prec_1$-minimal R-ancestor and S-ancestor of $(R[i], S[j])$ are called its R-parent and S-parent, see Fig. 5. Finally, we call $(R[i_1], R[j_1])$ and $(S[i_2], S[j_2])$ dependent if they have a common descendant and independent, otherwise.

Suppose $C' = (R', S', I')$ is a subcomplex induced by $R' = R[i_1, j_1]$ and $S' = S[i_2, j_2]$ and suppose furthermore there exists an exterior arc, $(R[a], S[b])$, with ancestors $(R[i], R[j])$ and $(S[i'], S[j'])$. The arc $(R[i], R[j])$ is C'-subsumed in $(S[i'], S[j'])$, if for any $(R[k], S[k']) \in I'$ with $i < k < j$, there exists some k' such that $i' < k' < j'$. In case of $C' = C$, we call $(R[i], R[j])$ simply "subsumed" in $(S[i'], S[j'])$, see Fig. 6. If $(R[i_1], R[j_1])$ is subsumed in $(S[i_2], S[j_2])$ and vice versa, we call these arcs equivalent.
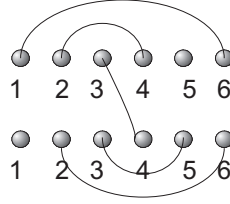
FIGURE 5. Ancestors and parents: for the exterior arc $(R[3], S[4])$, we have the following ancestor sets $A_R(R[3], S[4]) = \{(R[1], R[6]), (R[2], R[4])\}$ and $A_S(R[3], S[4]) = \{(S[2], S[6]), (S[3], S[5])\}$. In particular, $(R[2], R[4])$ and $(S[3], S[5])$ are the R-parent and S-parent respectively.
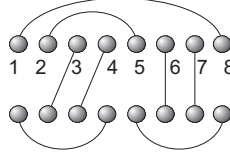

FIGURE 6. Subsumed and equivalent arcs: $(R[1], R[8])$ subsumes $(S[1], S[4])$ and $(S[5], S[8])$. Furthermore, $(R[2], R[5])$ is equivalent to $(S[1], S[4])$.

A joint structure, $J(R[i, j]; S[h, \ell], I') = J(i, j; h, \ell)$ is a subcomplex of $C(R, S, I)$ with the following properties, see Fig. 7:

• R, S are secondary structures

• there exist no external pseudoknots, i.e. if $(R[i_1], S[j_1]), (R[i_2], S[j_2]) \in I'$ where $i_1 < i_2$, then $j_1 < j_2$.

• there exist no "zig-zags", see Fig.8. I.e. if $(R[i_1], R[j_1])$ and $(S[i_2], S[j_2])$ are dependent, then either $(R[i_1], R[j_1])$ is subsumed by $(S[i_2], S[j_2])$ or vice versa. In absence of exterior arcs we
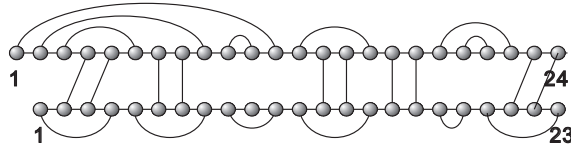

FIGURE 7. A joint structure induced by $R[1, 24]$ and $S[1, 23]$.

refer to a joint structure as a secondary structure segment, or segment for short. We call $S[i_1, j_1]$ maximal if there exists no segment, $S[i, j]$, containing $S[i_1, j_1]$. We remark that the idea of a joint structure goes back to [2] and has also been utilized in [11]. One key idea in our approach is to

introduce a specific joint structure, called a tight, which is in some sense a generalization of the loop. It can be viewed as the transitive closure of a loop with respect to exterior arcs.
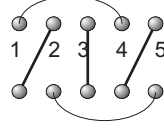

FIGURE 8. A zig-zag, generated by $(R[2], S[1])$, $(R[3], S[3])$ and $(R[5], S[4])$.

Let $J(a, b; c, d)$ be a fixed joint structure. A joint structure, $J(i, j; h, \ell) \subset J(a, b; c, d)$ is $J(a, b; c, d)$-tight (or tight in $J(a, b; c, d)$) if:
- there exists at least one exterior arc $(R[i_1], S[j_1])$
- for any $(R[i_1], S[j_1])$, we have

$$(2.3) \qquad (A_R(R[i_1], S[j_1]) \cup A_S(R[i_1], S[j_1])) \cap J(a, b; c, d) \in J(i, j; h, \ell)$$

- $J(i, j; h, \ell)$ is minimal with respect to $\subset$.

Given a tight (tjs), $J(i, j; h, \ell)$, we observe that neither one of the vertices i, j, h and ℓ, are start or endpoint of a segment. In particular, i, j, h and ℓ are not isolated. In combination with the non zig-zag property, we observe that there are only the following four types of tights ($\triangledown$), ($\triangle$), ($\square$) or ($\circ$), see Fig.9:
($\triangledown$): $(R[i], R[j]) \in J(i, j; h, \ell)$ and $(S[h], S[\ell]) \notin J(i, j; h, \ell)$
($\triangle$): $(S[h], S[\ell]) \in J(i, j; h, \ell)$ and $(R[i], R[j]) \notin J(i, j; h, \ell)$
($\square$): $\{(R[i], R[j]), (S[h], S[\ell])\} \in J(i, j; h, \ell)$
($\circ$): $\{(R[i], S[h])\} = J(i, j; h, \ell)$ and $i = j$, $h = \ell$, i.e. we have a single interaction.
Let $J_A(i, j; h, \ell)$ denote a tight structure $J(i, j; h, \ell)$ having type ξ, where $\xi \in A \subset \{\triangledown, \triangle, \square, \circ\}$. In particular, $J_\xi(i, j; h, \ell)$ is a tight structure $J(i, j; h, \ell)$ of type ξ.
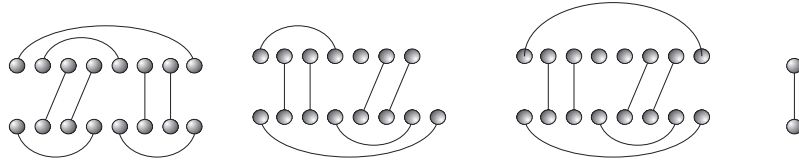

FIGURE 9. From left to right: (tjs) of type ($\triangledown$), ($\triangle$), ($\square$) and ($\circ$).

Proposition 2.1. *Let $J(a, b, c, d)$ be a joint structure, then the following assertions hold:*

(a) *if $J(i, j; h, \ell)$ is tight in $J(a, b, c, d)$, then $J(i, j; h, \ell)$ has type $\tau \in \{\triangledown, \triangle, \square, \circ\}$*

(b) *any exterior arc is contained in a unique $J(a, b, c, d)$-(tjs)*

(c) *$J(a, b, c, d)$ decomposes into a unique sequence of (tjs) and maximal segments.*

Suppose we are given two exterior arcs $(R[i_1], S[j_1]), (R[i_2], S[j_2]) \in J(i, j; h, \ell)$. For two $J(i, j; h, \ell)$-tight structures, $J_T((R[i_1], S[j_1])), J_T((R[i_2], S[j_2]))$ we set

$$J_T((R[i_1], S[j_1])) = J_T((R[i_2], S[j_2])) \iff (R[i_1], S[j_1])_1 \sim_{J(i,j;h,\ell)} (R[i_2], S[j_2]).$$

Suppose $J_T(i, j; r, s)$ is a tight structure where $i \leq a < b \leq j$ and $r \leq c < d \leq s$. A double-tight structure $J_{DT}(i, j; r, s)$ in $J_T(i, j; r, s)$, is a joint structure $J(i, j; r, s)$ such that $J(i, j; r, s) \subset J_T(i, j; r, s)$ and

$$(2.4) \qquad J_{DT}(i, j; r, s) = (J_T(i, a; r, c), J(a+1, b-1; c+1, d-1), J_T(b, j; d, s))$$

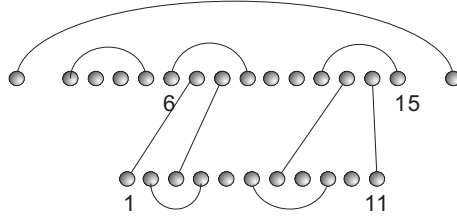where $J_T(i, a; r, c)$ and $J_T(b, j; d, s)$ are $J(a+1, b-1; c+1, d-1)$-tight structures, see Fig. 10.

FIGURE 10. A double-tight structure $J_{DT}(6, 15; 1, 11)$ in $J(2, 15; 1, 11)$. Note that the joint structure $J(1, 15; 1, 11)$ itself is a $\triangledown$-tight.
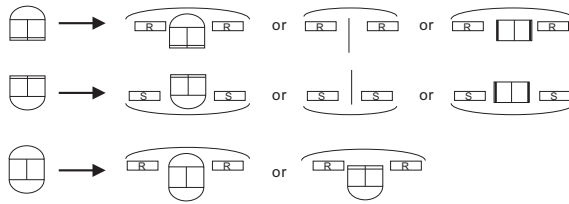

FIGURE 11. Decomposing tights: we show how to decompose a tights of the types $(\triangledown)$, $(\triangle)$ or $(\square)$ via to Corollary 2.2, Corollary 2.3 and Corollary 2.4.

Corollary 2.2. *Let $J_\bigtriangledown(i,j;r,s)$ be a tight structure of type $\bigtriangledown$ and let $\zeta_1 = (R[h_1], S[\ell_1])$ and $\zeta_2 = (R[h_2], S[\ell_2])$ be the minimal and maximal exterior arcs in $J(i,j;r,s)$ and $i+1 \le i_1 \le j_1 \le j-1$. Then*

(2.5)
$$J(i+1, j-1; r, s) =$$

$$\begin{cases} (R[i+1, i_1-1], J_{\{\bigtriangledown,\circ\}}(i_1, j_1; r, s), R[j_1+1, j-1]), & \text{if } \zeta_1 \sim_{J(i+1,j-1;r,s)} \zeta_2 \\ (R[i+1, i_1-1], J_{DT}(i_1, j_1; r, s), R[j_1+1, j-1]) & \text{otherwise,} \end{cases}$$

where $J_{\{\bigtriangledown,\circ\}}(i_1, j_1; r, s)$ denotes a $J(i+1, j-1; r, s)$-tight of type $\bigtriangleup$ or $\circ$.

Of course we have

Corollary 2.3. *Let $J_\bigtriangleup(i,j;r,s)$ be a tight structure of type $\bigtriangleup$ and let $\zeta_1 = (R[h_1], S[\ell_1])$ and $\zeta_2 = (R[h_2], S[\ell_2])$ be the minimal and maximal exterior arcs in $J(i,j;r,s)$ and $r+1 \le r_1 \le s_1 \le s-1$. Then*

(2.6)
$$J(i, j; r+1, s-1) =$$

$$\begin{cases} (S[r+1, r_1-1], J_{\{\bigtriangleup,\circ\}}(i, j; r_1, s_1), S[s_1+1, s-1]), & \text{if } \zeta_1 \sim_{J(i,j;r+1,s-1)} \zeta_2 \\ (S[r+1, r_1-1], J_{DT}(i, j; r_1, s_1), S[s_1+1, s-1]), & \text{otherwise,} \end{cases}$$

where $J_{\{\bigtriangleup,\circ\}}(i_1, j_1; r, s)$ denotes a $J(i, j; r+1, s-1)$-tight of type $\bigtriangleup$ or $\circ$.

Corollary 2.4. *Let $J(i,j;r,s)$ be a tight structure of type $\square$ and set $i+1 \le i_1 \le j_1 \le j-1$, then $J_t(i,j;r,s)$ decomposes as follows:*

(2.7) $$J(i+1, j-1; r, s) = (R[i+1, i_1-1], J_{\{\bigtriangleup,\square\}}(i_1, j_1; r, s), R[j_1+1, j-1]),$$

where $J_{\{\bigtriangleup,\square\}}(i_1, j_1; r, s)$ denotes a $J(i+1, j-1; r, s)$-tight of type $\bigtriangleup$ or $\square$.

2.1. **Proofs. Proof of Proposition 2.1**

Proof. Let $(R[i], S[j])$ be the maximal (rightmost) exterior arc of $J(a,b,c,d)$. We consider the set of maximal $(R[i], S[j])$-ancestors, M. In case of $M = \varnothing$ we immediately observe $J(i,j;h,\ell) = (R[i], S[j])$, i.e. $J(i,j;h,\ell)$ is of type $\circ$. Suppose next $|M| = 1$. By symmetry we can, without loss of generality, assume $M = \{(R[i_1], R[j_1])\}$. Let $(R[i_0], S[j_0])$ the minimal exterior arc being

an descendant of $(R[i_1], R[j_1])$ and let j_0^* denote either the startpoint of the maximal $(R[i_0], S[j_0])$ S-ancestor or set $j_0^* = j_0$ if no such ancestor exists. Then, by construction, $J(i_1, j_1; j_0^*, j)$ is tight in $J(a, b, c, d)$. Finally, in case of $|M| = 2$, i.e. $M = \{(R[i_1], R[j_1]), (S[r_1], S[s_1])\}$. We may, without loss of generality, assume that $(R[i_1], R[j_1])$ subsumes $(S[r_1], S[s_1])$. Again we consider the minimal descendant of $(R[i_1], R[j_1])$, $(R[z], S[x])$. Let x^* be either the startpoint of the maximal S-ancestor of $(R[z], S[x])$ or $x^* = x$, otherwise. Then $J(i_1, j_1; x^*, s_1)$ is tight. If $(R[i_1], R[j_1])$ is equivalent to $S[r_1], S[s_1])$, then $J(i_1, j_1; r_1, s_1)$ is tight. In the above procedure we have constructed a (tjs), J^*, of type $\tau \in \{\triangledown, \triangle, \square, \circ\}$ that contains the maximal exterior $J(a, b, c, d)$-arc. By definition of tight and the fact that we have noncrossing arcs it follows that any other (tjs) of $J(a, b, c, d)$ is disjoint to J^*. We proceed by considering the rightmost exterior arc of $J(a, b, c, d)$ that is not contained in J^*, concluding assertion (c) by induction on the number of exterior arcs of $J(a, b, c, d)$. Since any exterior arc of $J(a, b, c, d)$ is contained in a unique (tjs) generated by the above procedure, (b) follows, see Fig. 12. $\square$
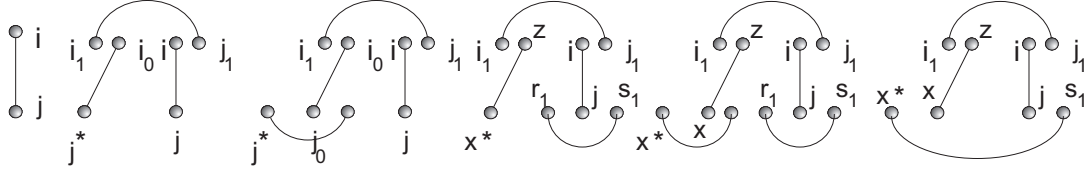

FIGURE 12. Illustration of Prop.2.1.

Proof to Corollary 2.2

Proof. According to Prop. 2.1(b), there exist unique $J(i+1, j-1; r, s)$-tight structures $J(i_1, i_2; r, r_1)$ and $J(j_2, j_1; s_1, s)$ such that $J(i_1, i_2; r, r_1) = J_T(\zeta_1)$ and $J(j_2, j_1; s_1, s) = J_T(\zeta_2)$, respectively. We have the following two scenarios: in case of $\zeta_1 \sim_{J(i+1, j-1; r, s)} \zeta_2$, i.e. $J_T(\zeta_1) = J_T(\zeta_2)$, we have either $r = s$, in which case $J(i_1, j_1; r, s)$ is of type $\circ$ and in view of $(S[r], S[s]) \notin J_{\triangledown}(i, j; r, s)$ $J(i_1, j_1; r, s)$ is of type $\triangledown$, otherwise. In case of $\zeta_1 \nsim_{J(i+1, j-1; r, s)} \zeta_2$, $J(i_1, j_1; r, s)$ is a $J(i+1, j-1; r, s)$-double tight structure. $\square$

Proof of Corollary 2.4

Proof. We observe that there exist only one $J(i+1, j-1; r, s)$-tight structure, since $(S[r], S[s]) \in J(i+1, j-1; r, s)$. We consider the set M, consisting of arcs that are equivalent to $(S[r], S[s])$. According to Prop. 2.1, (c), we have

$$J(i+1, j-1; r, s) = \begin{cases} (R[i+1, i_1-1], J_\triangle(i_1, j_1; r, s), R[j_1+1, j-1]) & \text{for} M = \varnothing \\ (R[i+1, i_1-1], J_\square(i_1, j_1; r, s), R[j_1+1, j-1]) & \text{otherwise.} \end{cases}$$

$\square$

3. Unique decomposition

We showed in Section 2 via Prop. 2.1 that an arbitrary joint structure uniquely decomposes into a sequence of segments and tight structures. Via the combinatorial corollaries, Cor. 2.2, Cor. 2.3 and Cor. 2.4 we introduced a unique decomposition procedure for tights, see Fig. 13 and Fig. 14, below.
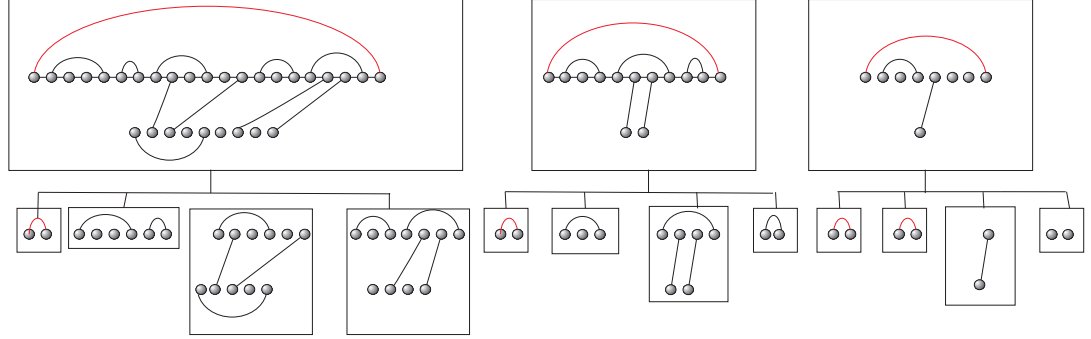

Figure 13. Illustration of Cor. 2.2.

In this section we give the algorithmic interpretation of the above results. In the course of our analysis we derive for any joint structure $J(1, N; 1, M)$ a unique decomposition tree via Procedure (a), (b) and (c), below, see Fig. 15. Let us begin by giving an interpretation of Prop. 2.1.

Procedure (a):

input: a joint structure $\vartheta_0 = J(i, j; h, \ell)$, which is not ϑ_0-tight or a ms

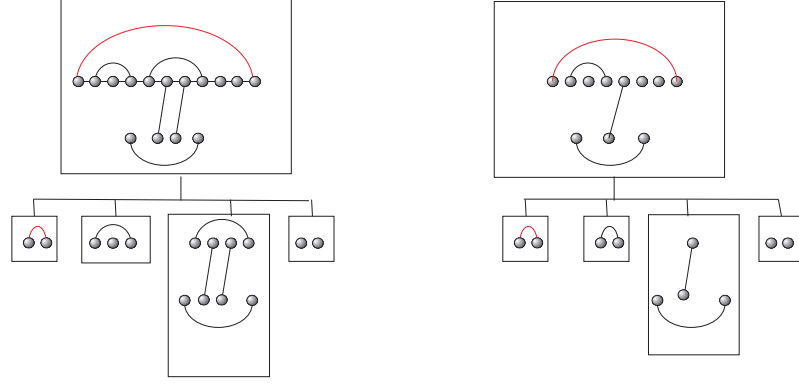output: a unique tree $T_a(\vartheta_0) = (V_a(T), E_a(T))$

FIGURE 14. Illustration of Cor. 2.4.
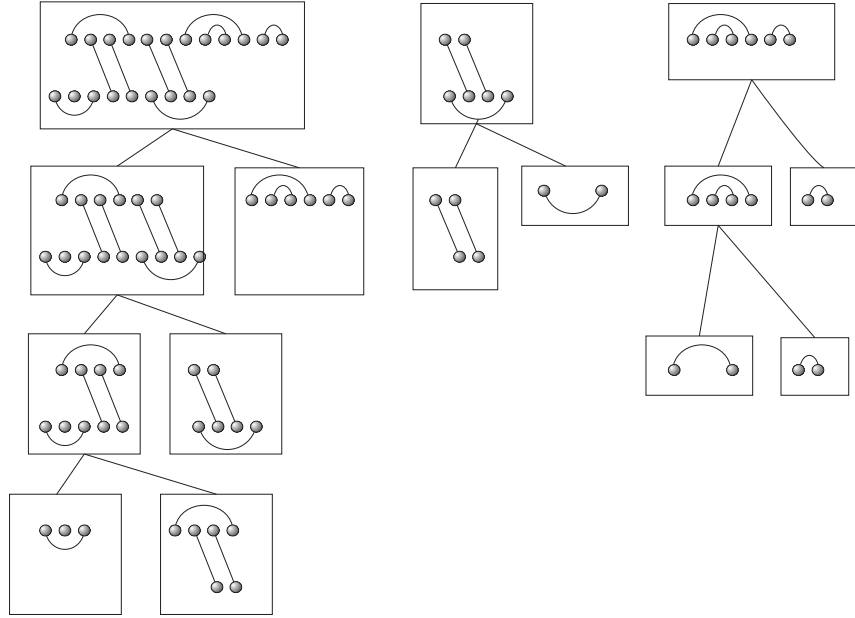
FIGURE 15. Illustration of Procedure (a), Procedure (b) and Procedure (c) for the joint structure $J(1, 12, 1, 8)$. From left to right we display $T_a(1, 12; 1, 8)$, $T_b(5, 6; 6, 9)$ and $T_c(R[7, 12])$.

Let $i \leq j^* \leq j + 1$ and $R[j^*, j]$ be the ϑ_0-ms contain j. In particular, $j^* = j + 1$ in case of such an ms does not exist and $j^* = 1$ if $R[i, j]$ itself is a ms. Analogously, we define $S[\ell^*, \ell]$. We construct

the tree $T_a(\vartheta_0)$ recursively as follows:

initialization: $V_a(T) = \{\vartheta_0\}$ and $E_a(T) = \varnothing$.

(a1): in case of $j^* = j+1$ and $\ell^* = \ell+1$, i.e. ϑ_0 is right-tight, then ϑ_0 decomposes via Prop. 2.1 (b) and (c) into a ϑ_0-tight structure $\vartheta_1 = J_{\{\bigtriangledown,\triangle,\square,\circ\}}(i_1, j; h_1, \ell)$ and a joint structure $\vartheta_2 = J(i, i_1 - 1; h, h_1 - 1)$, where $i \le i_1 \le j$ and $h \le h_1 \le \ell$. Accordingly, we have

$$(3.1) \qquad V_a(T) \;=\; V_a(T) \cup \{\vartheta_1, \vartheta_2\},$$

$$(3.2) \qquad E_a(T) \;=\; E_a(T) \cup \{(\vartheta_0, \vartheta_1), (\vartheta_0, \vartheta_2)\}.$$

(a2) otherwise, ϑ_0 decomposes into a ϑ_0-right tight structure $\vartheta_3 = J_{RT}(i, j^* - 1; h, \ell^* - 1)$ and two ms $\vartheta_4 = R[j^*, j]$, $\vartheta_5 = S[\ell^*, \ell]$. Accordingly, we have

$$(3.3) \qquad V_a(T) \;=\; V_a(T) \cup \{\vartheta_3, \vartheta_4, \vartheta_5\},$$

$$(3.4) \qquad E_a(T) \;=\; E_a(T) \cup \{(\vartheta_0, \vartheta_3), (\vartheta_0, \vartheta_4), (\vartheta_0, \vartheta_5)\}.$$

We iterate the process until all the leaves of $T_a(\vartheta_0)$ are either ϑ_0-tight structures or ϑ_0-ms.

We proceed by providing an interpretation of Cor. 2.2, Cor. 2.3 and Cor. 2.4.

Procedure (b):

input: a tight structure $\vartheta_0 = J(i, j; h, \ell)$

output: a unique tree $T_b(\vartheta_0) = (V_b(T), E_b(T))$

initialization: $V_b(T) = \{\vartheta_0\}$ and $E_b(T) = \varnothing$.

We distinguish $J(i, j; h, \ell)$ by type:

$\circ$: do nothing.

$\square$: according to Cor. 2.4, ϑ_0 decomposes into $\vartheta_1 = (R[a], R[b])$, $\vartheta_2 = R[i + 1, i_1 - 1]$, $\vartheta_3 = J_{\square,\triangle}(i_1, j_1; h, \ell)$ and $\vartheta_4 = R[j_1 + 1, j - 1]$, which gives rise to

$$(3.5) \qquad V_b(T) \;=\; V_a(T) \cup \{\vartheta_1, \vartheta_2, \vartheta_3 \vartheta_4, \vartheta_5\},$$

$$(3.6) \qquad E_b(T) \;=\; E_a(T) \cup \{(\vartheta_0, \vartheta_1), (\vartheta_0, \vartheta_2), (\vartheta_0, \vartheta_3), (\vartheta_0, \vartheta_4), (\vartheta_0, \vartheta_5)\}.$$

$\bigtriangledown$: according to Cor 2.2, we consider the set of $J(i + 1, j - 1; h, \ell)$-tight structures, denoted by M. In case of $|M| = 1$, $J(i + 1, j - 1; h, \ell)$ decompose into a sequence of a $J(i + 1, j - 1; h, \ell)$-tight structure $\vartheta_6 = J_{\{\bigtriangledown,\circ\}}(i + 1, j - 1; h, \ell)$ and two $J(i + 1, j - 1; h, \ell)$-ms, $\vartheta_7 = R[i + 1, i_1 - 1]$ and $\vartheta_8 = R[j_1 + 1, j - 1]$, where $i \le i_1 < j_1 \le j$. Accordingly,

$$(3.7) \qquad V_b(T) \;=\; V_a(T) \cup \{\vartheta_1, \vartheta_6, \vartheta_7, \vartheta_8\},$$

$$(3.8) \qquad E_b(T) \;=\; E_a(T) \cup \{(\vartheta_0, \vartheta_1), (\vartheta_0, \vartheta_6), (\vartheta_0, \vartheta_7), (\vartheta_0, \vartheta_8)\}.$$

In case of $|M| > 1$, $J(i+1, j-1; h, \ell)$ decomposes into a sequence consisting of a $J(i+1, j-1; h, \ell)$-double tight structure $\vartheta_9 = J_{DT}(i+1, j-1; h, \ell)$ and two $J(i+1, j-1; h, \ell)$-ms. $\vartheta_7 = R[i+1, i_1-1]$ and $\vartheta_8 = R[j_1 + 1, j - 1]$, where $i \leq i_1 < j_1 \leq j$. Accordingly,

$$(3.9) \qquad V_b(T) = V_a(T) \cup \{\vartheta_1, \vartheta_7, \vartheta_8, \vartheta_9\},$$

$$(3.10) \qquad E_b(T) = E_a(T) \cup \{(\vartheta_0, \vartheta_1), (\vartheta_0, \vartheta_7), (\vartheta_0, \vartheta_8), (\vartheta_0, \vartheta_9)\}.$$

Furthermore, let $i_1 \leq i_2 < j_1$ and $h \leq j_2 < \ell$, a $J(i+1, j-1; h, \ell)$-double tight structure $\vartheta_9 = J_{DT}(i+1, j-1; h, \ell)$ decomposes into a $J(i+1, j-1; h, \ell)$-tight structure $\vartheta_{10} = J_{\{\triangledown, \circ, \triangle, \square\}}(i_1, i_2; h, j_2)$ and a $J(i+1, j-1; h, \ell)$-right tight structure $\vartheta_{11} = J_{RT}(i_2 + 1, j_1; j_2 + 1, \ell)$. I.e.

$$(3.11) \qquad V_b(T) = V_a(T) \cup \{\vartheta_{10}, \vartheta_{11}\},$$

$$(3.12) \qquad E_b(T) = E_a(T) \cup \{(\vartheta_9, \vartheta_{10}), (\vartheta_9, \vartheta_{11})\}.$$

$\triangle$: analogous to type $\triangledown$ via symmetry.
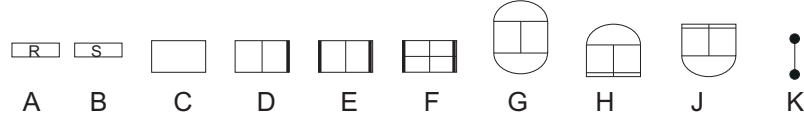In Fig. 17 we give an overview of Procedure (a) and Procedure (b).

FIGURE 16. (A), (B): maximal secondary segments (ms) $R[i, j]$, $S[r, s]$, (C): joint structure $J(i, j; r, s)$, (D) right-tight structures $J_{RT}(i, j; r, s)$, (E): double-tight structure $J_{DT}(i, j; r, s)$, (F): $J_{\triangledown}(i, j; r, s)$, a tight structure of type $\triangledown$, $\triangle$ or $\square$, (G): $J_{\square}(i, j; r, s)$, (H): $J_{\triangledown}(i, j; r, s)$, (J): $J_{\triangle}(i, j; r, s)$ and (K): exterior arc.

Finally, we have the wellknown [32] secondary structure loop-decomposition
Procedure (c)**:**
input: a secondary structure $\vartheta_0 = R[i, j]$
output: a tree $T_c(\vartheta_0) = (V_c(T), E_c(T))$
initialization: $V_b(T) = \{\vartheta_0\}$ and $E_b(T) = \varnothing$.
We distinguish the following two cases:
$(c1)$: in case of $(R[i], R[j]) \notin R[i, j]$, let $\varnothing_a^b$ denote empty segment in which all the vertices are isolated. For $1 \leq j^* \leq j + 1$, let $\varnothing_{j^*}^j$ be the maximal empty segment that contains $R[j]$. In particular, if j is not isolated, we have $j^* = j + 1$. Let $R^b[i_1, j^* - 1]$ denote the segment in which $R[i_1]$ is connected with $R[j^* - 1]$. Then $R[i, j]$ decomposes as follows $R[i, j] = (\vartheta_1 = R[i, i_1 - 1], \vartheta_2 =$
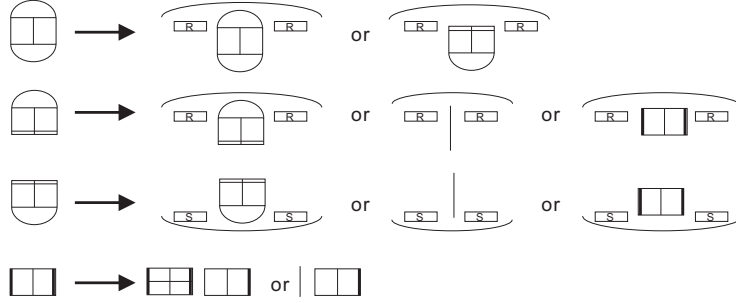
Procedure (a)
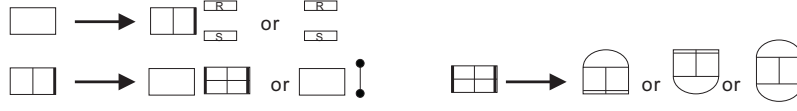
Procedure (b)

FIGURE 17. Illustration of Procedure (a) and Procedure (b), notations are given via Fig. 16 above.

$R^b[i_1, j^* - 1], \vartheta_3 = \varnothing_{j^*}^j)$ and we set

$$(3.13) \qquad\qquad V_c(T) \;=\; V_c(T) \cup \{\vartheta_1, \vartheta_2, \vartheta_3\},$$

$$(3.14) \qquad\qquad E_c(T) \;=\; E_c(T) \cup \{(\vartheta_0, \vartheta_1), (\vartheta_0, \vartheta_2), (\vartheta_0, \vartheta_3)\}.$$

$(c2)$: in case of $(R[i], R[j]) \in R[i, j]$, i.e. for $R[i, j] = R^b[i, j]$, we have a decomposition into the pair $(\vartheta_4 = (R[i], R[j]), \vartheta_5 = R[a + 1, b - 1])$. Accordingly, we have $V_c(T) = V_c(T) \cup \{\vartheta_4, \vartheta_5\}$ and $E_c(T) = E_c(T) \cup \{(\vartheta_0, \vartheta_4), (\vartheta_0, \vartheta_5)\}$.
We iterate (c1) and (c2), until all the leaves in T are either isolated segments or single arcs.

For any joint structure, $J(1, N; 1, M)$, we can now construct a tree, with root $J(1, N; 1, M)$ and whose vertices are specific subgraphs of $J(1, N; 1, M)$. The latter are obtained by successive application of Procedure (a), (b) and (c), see Fig. 28. To be precise, let H be the graph rooted in $J(1, N; 1, M)$ defined inductively as follows: for the induction basis for fixed $J(1, N; 1, M)$ only one, Procedure (a), (b) or (c) applies. Procedure (a), (b) or (c) generates the (procedure-specific, nontrivial) subtrees, T_a, T_b and T_c. Suppose $\vartheta_\dagger$ is a leaf of T that has been constructed via Procedure (a), (b) or (c). As in case of the induction basis, each such leaf is input for exactly one procedure, which in turn generates a corresponding subtree. Prop. 2.1, Cor. 2.2, Cor. 2.3 and

Cor. 2.4 imply that H itself is a tree. We denote this decomposition tree by $T(1, N; 1, M)$, see Fig. 28. Accordingly, we have proved

Observation 1. *For any joint structure, $J(1, N; 1, M)$, there exists a unique decomposition tree, $T(1, N; 1, M)$, whose leafs are either interior or exterior $J(1, N; 1, M)$-arcs or isolated segments.*

As we shall see in Section 5, the decomposition tree plays a key role for the calculation of the base pairing probabilities. To be precise, given a joint structure, $J(i, j; h, \ell)$, let $T_J(1, N; 1, M)$ be the decomposition tree of $J(1, N; 1, M)$ and let $\Sigma_0 = \{J(1, N; 1, M) \mid J(i, j; h, \ell) \in T_J(1, N; 1, M)\}$. Then the probability of $J(i, j; h, \ell)$, denoted by $\mathbb{P}(i, j; h, \ell)$, is given by

$$(3.15) \qquad \mathbb{P}(i, j; h, \ell) = \sum_{J(1, N; 1, M) \in \Sigma_0} \mathbb{P}(J(1, N; 1, M))$$

and furthermore

Observation 2. *In general $J(i, j; h, \ell) \subset J(1, N; 1, M)$ is not equivalent to $J(i, j; h, \ell) \in T_J(1, N; 1, M)$, see Fig. 18. However, in case of secondary structures, i.e. $J(i, j; h, \ell) = (R[i], R[j])$, we have*

$$(3.16) \qquad (R[i], R[j]) \subset J(1, N; 1, M) \Leftrightarrow (R[i], R[j]) \in T_J(1, N; 1, M).$$
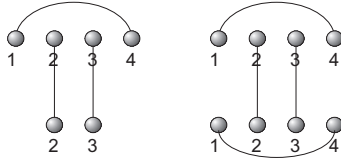
FIGURE 18. $J(1, 4; 2, 3)$ has the property that $J(1, 4; 2, 3) \subset J(1, 4; 1, 4)$ but $J(1, 4; 2, 3) \notin T_J(1, 4; 1, 4)$.

4. From the decomposition tree to the partition function

We discussed in the introduction the concept of the loop-based partition function of RNA secondary structures due to McCaskill [20]. We observed there that the key property for its derivation is the unique decomposition into substructures and their recursive analysis. For instance, suppose we

are given a tight of type $\triangledown$ from which we remove, by virtue of Cor. 2.2, its outer arc. For this purpose, the context of the latter, i.e. its particular arc-configuration has to be taken into account. However, once the unique decomposition is established, the existence of specific subclasses of joint structures allowing for the dynamic programming of the partition function follows. We remark that the particular choice of the latter may not be unique.

The first step is to extend the standard loop-energy model for secondary structures by introducing two new loop-types due to Chitsaz *et al.* [11]: the kissing loop and the hybrid, see Figure 19.

4.1. **Loops.** Having discussed the standard loop types of secondary structures in Section 1, we proceed now by introducing the loops that contain exterior arcs.

(4) a *hybrid*-loop (Hy)is a sequence $((R[i_1], S[j_1]), \ldots, ([R[i_s], S[j_s]))$, where $s \geq 2$ and (i_r, j_r) is nested in (i_1, j_1) such that $R[i_h + 1, i_{h+1} - 1] = \varnothing_{i_h+1}^{i_{h+1}-1}$ and $S[j_h + 1, j_{h+1} - 1] = \varnothing_{j_h+1}^{j_{h+1}-1}$.

(5) a *kissing*-loop (K) is either a pair, $((R[i], R[j]), R[i + 1, j - 1])$, such that there exists at least one $(R[i], R[j])$-child, $(R[i_1], S[j_1])$ where $i < i_1 < j$ or a pair $((S[i], S[j]), S[i + 1, j - 1])$, with $(R[i], R[j])$-child $(R[i_1], S[j_1])$ and $i < j_1 < j$.
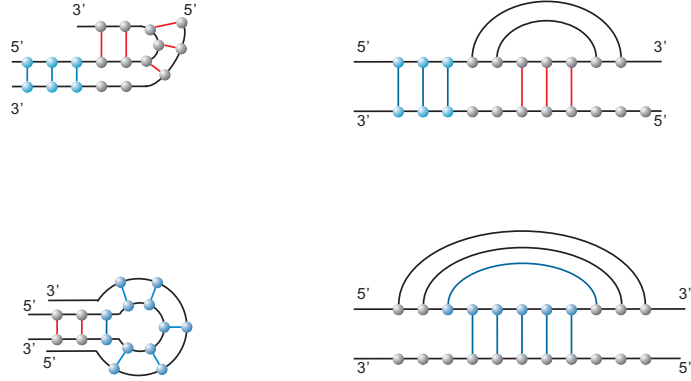
FIGURE 19. The two new loop types: the hybrid (top) and the kissing loop (bottom).

The arguments of Prop. 2.1, Cor. 2.2, Cor. 2.3 and Cor. 2.4 imply that each joint structure can uniquely be decomposed into a sequence of loops–a necessary and sufficient condition for the mfe-folding of joint structures. As we shall see in the next section, the unique decomposition *and* the particular choice of loops give rise to specific subclasses via which the partition function can be

recursively expressed. Furthermore, following [7], we allow for an initiation energy, i.e. each hybrid loop is given an energy penalty of σ_0. In addition, we allow for a scaling, $0 < \sigma \leq 1$, of the energy contribution of each hybrid loop. As default we set $\sigma_0 = 0$, $\sigma = 1$.

4.2. **Case studies.** Consider a joint structure $J(i, j; h, \ell) \in T(J(1, N; 1, M))$. For the purpose of assigning an energy to a substructure, we have to distinguish substructures by their "outer" loop type, see Case 1 as well as Fig. 2 and Fig 19. To convey the key ideas we shall restrict our analysis to three case studies.
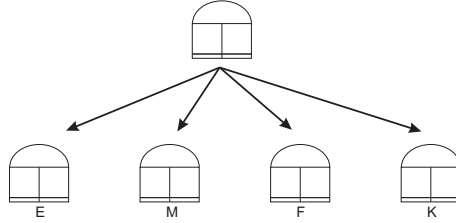

FIGURE 20. Context dependency: the labels "$\mathsf{E}, \mathsf{M}, \mathsf{F}, \mathsf{K}$" are defined in Case 1. We display from left to right $J_{\triangledown}^{\mathsf{E}}(i, j; h, \ell)$, $J_{\triangledown}^{\mathsf{M}}(i, j; h, \ell)$, $J_{\triangledown}^{\mathsf{F}}(i, j; h, \ell)$ and $J_{\triangledown}^{\mathsf{K}}(i, j; h, \ell)$, respectively.

Given a joint structure $J(i, j; h, \ell) \in T(J(1, N; 1, M))$, we set $M_R(i, j) = \{(R[i_1], R[j_1]) \mid i_1 < i \leq j < j_1\}$ and $M_S(h, \ell) = \{(S[i_1], S[j_1]) \mid i_1 < h \leq \ell < j_1\}$.

Case 1. Suppose we are given a tight structure $J_{\triangledown}(i, j; h, \ell)$. In case of $M_S(h, \ell) = \varnothing$, we call $S[h, \ell]$ external and use the notation $J_{\triangledown}^{\mathsf{E}}(i, j; h, \ell)$. Otherwise, let $(S[i_0], S[j_0])$ be the minimal element of $M_S(h, \ell)$. We denote the type of the loop including $(S[i_0], S[j_0])$, by ξ. In case of $\xi = \mathsf{M}$, we use the notation $J_{\triangledown}^{\mathsf{M}}(i, j; h, \ell)$. Otherwise, in case of $\xi = \mathsf{K}$, we write $J_{\triangledown}^{\mathsf{K}}(i, j; h, \ell)$ or $J_{\triangledown}^{\mathsf{F}}(i, j; h, \ell)$ depending on whether or not $J_{\triangledown}(i, j; h, \ell)$ contains the child of $(S[i_0], S[j_0])$, see Fig. 20.

Case 2. Suppose we are given a double-tight structure, $J_{DT}(i, j; h, \ell)$. Then we arrive at the twelve subclasses presented in Figure 21. Indeed, according to Cor. 2.2, there does not exist any $J_{DT}^{\mathsf{E},\mathsf{E}}(i, j; h, \ell)$, i.e. $M_R(i, j) \cup M_S(h, \ell) \neq \varnothing$. Without loss of generality, we may assume that $M_R(i, j) \neq \varnothing$ and that $(R[i_1], R[j_1]) \in M_R(i, j)$ is minimal. In case of $M_S(h, \ell) = \varnothing$, we use the notation $J_{DT}^{\mathsf{Y},\mathsf{E}}(i, j; h, \ell)$, where Y is the loop type formed by $(R[i_1], R[j_1])$ and $R[i_1 + 1, j_1 - 1]$.
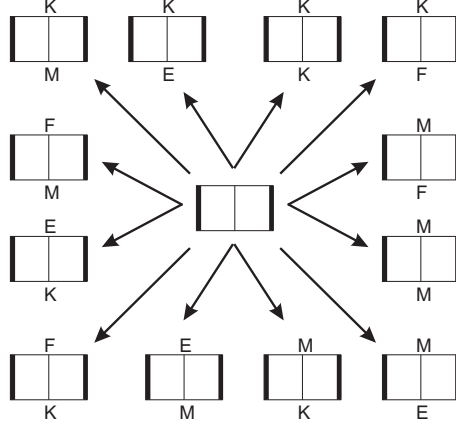
FIGURE 21. The twelve subclasses of $J_{DT}(i,j;h,\ell)$ as discussed in Case 2.

Otherwise, we have $M_S(h,\ell) \neq \varnothing$. Let $(S[i_2], S[j_2])$ be the minimal element. In this case we use the notation $J_{DT}^{Y_1,Y_2}(i,j;h,\ell)$, where Y_1 and Y_2 are the loop-types formed by $(R[i_1], R[j_1])$, $R[i_1+1, j_1-1]$ and $(S[i_2], S[j_2])$, $S[i_2+1, j_2-1]$, respectively, see Fig. 21.

*Case* 3. In case of a right-tight structure, $J_{RT}^{\mathsf{K,K}}(i,j;h,\ell)$, we obtain four subclasses. In case of $(R[j], S[\ell]) \in J_{RT}^{\mathsf{K,K}}(i,j;h,\ell)$, we say $J_{RT}^{\mathsf{K,K}}(i,j;h,\ell)$ is (rB) and (rA), otherwise. Let $(R[i_1], S[j_1])$ denote the minimal exterior arc in $J_{RT}^{\mathsf{K,K}}(i,j;h,\ell)$. According to Prop. 2.1, there exists a unique $J(i,j;h,\ell)$-tight structure $J_T(R[i_1], S[j_1])$, such that $(R[i_1], S[j_1]) \in J_T(R[i_1], S[j_1])$. In case of $J_T(R[i_1], S[j_1])$ is of type $\circ$, i.e. $(R[i_1], S[j_1])$ itself and $R[i, i_1] = \varnothing_i^{i_1}$, $S[h, j_1] = \varnothing_h^{j_1}$, we say $J_{RT}^{\mathsf{K,K}}(i,j;h,\ell)$ is (lB) and (lA), otherwise. We use the notation $J_{RT}^{\mathsf{K,K},Y_1,Y_2}(i,j;h,\ell)$, if $J_{RT}^{\mathsf{K,K}}(i,j;h,\ell)$ is (lY_1) and (rY_2), respectively, see Fig. 22.

4.3. **The partition function.** In the previous section we discussed specific subclasses of joint structures. They were designed to facilitate the recursive construction of the partition function. The purpose of this section is to showcase the respective recursions induced by these classes.

Case 1: $J_{\triangledown}^{\mathsf{M}}(i,j;r,s)$. According to Cor. 2.2, we have three cases: $J_{\triangledown}(i,j;r,s)$ decomposes into either a $J(i-1,j+1;r,s)$-tight structure of type κ, where $\kappa \in \{\triangledown, \circ\}$ or a $J(i-1,j+1;r,s)$-double tight structure and a ms. By definition of $J_{\triangledown}^{\mathsf{M}}(i,j;r,s)$, the case of a $J(i-1,j+1;r,s)$-tight structure of type $\circ$ is impossible. Considering the type of the loop including $(R[i], R[j])$ and $R[i+1, j-1]$,
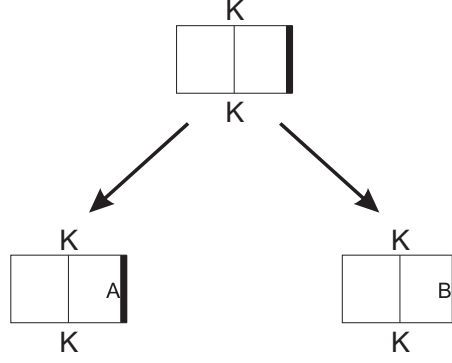
FIGURE 22. The four subclasses of $J_{RT}^{\mathsf{K},\mathsf{K}}(i,j;h,\ell)$, see Case 3 for details.

we arrive exactly at the four cases, denoted by I_1, I_2, I_3 and I_4, from left to right, displayed in Fig. 23.
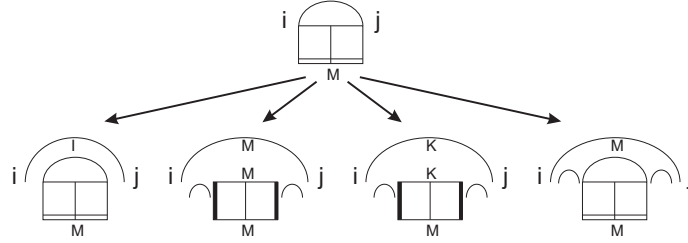
FIGURE 23. The four decompositions of $J_{\triangledown}^{\mathsf{M}}(i,j;r,s)$ via Procedure (b), denoted by I_1, I_2, I_3 and I_4, from left to right, respectively.

Let $i < h < \ell < j$. According to the recurrences displayed in Fig. 23, the partition function satisfies for $J_{\triangledown}^{\mathsf{M}}(i,j;r,s)$ the following recursion:

$$(4.1) \qquad Q_{\triangledown}^{\mathsf{M}}(i,j;r,s) = \sum_{h,\ell}(Q(I_1) + Q(I_2) + Q(I_3) + Q(I_4)),$$

where

$$\begin{aligned}
Q(I_1) &= Q_{\triangledown}^{\mathsf{M}}(h,\ell;r,s)e^{-\mathsf{Int}(i,j;h,\ell)/kT}, \\
Q(I_2) &= Q_{DT}^{\mathsf{M,M}}(h,\ell;r,s)e^{-(\alpha_1+\alpha_2)/kT}(e^{-(j-\ell-1)\alpha_3/kT} + Q^{\mathsf{m}}(i+1,h-1)) \\
&\quad \times(e^{-(h-i-1)\alpha_3/kT} + Q^{\mathsf{m}}(\ell+1,j-1)), \\
Q(I_3) &= Q_{DT}^{\mathsf{K,M}}(h,\ell;r,s)e^{-(\beta_1+\beta_2)/kT}(e^{-(j-\ell-1)\beta_3/kT} + Q^{\mathsf{k}}(i+1,h-1)) \\
&\quad \times(e^{-(h-i-1)\beta_3/kT} + Q^{\mathsf{k}}(\ell+1,j-1)), \\
Q(I_4) &= Q_{\triangledown}^{\mathsf{M}}(h,\ell;r,s)e^{-(\alpha_1+2\alpha_2)/kT}(Q^{-(j-\ell-1)\alpha_3/kT}Q^{\mathsf{m}}(i+1,h-1) \\
&\quad +e^{-(h-i-1)\alpha_3/kT}Q^{\mathsf{m}}(\ell+1,j-1)) + Q^{\mathsf{m}}(\ell+1,j-1)Q^{\mathsf{m}}(i+1,h-1)).
\end{aligned}$$

Case 2: $J_{DT}^{\mathsf{K,M}}(i,j;h,\ell)$. According to Procedure (b), a double tight structure decomposes into a $J(i,j;h,\ell)$-tight structure, $J(i,i_1;h,h_1)$ and a $J(i,j;h,\ell)$-right tight structure, $J(i_1+1,j;h_1+1,\ell)$. We observe that the type of the outer loop of $S[h,h_1]$ and $S[h_1+1,\ell]$ coincides with that of $S[h,\ell]$, i.e. M. Analogously, the outer loop of $R[i,i_1]$ and $R[i_1+1,j]$, denoted by $(R[i_0,j_0])$, is of type K. Furthermore, at least one of the substructures $R[i,i_1]$ and $R[i_1+1,j]$ contain the child of $(R[i_0,j_0])$. Consequently we arrive at the three scenarios labeled by from left to right by J_1, J_2 and J_3 displayed in Fig. 24. Setting
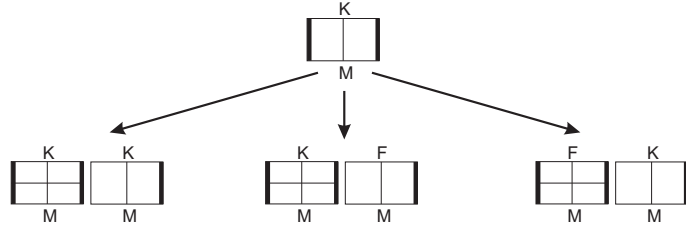

FIGURE 24. The decomposition of $J_{DT}^{K,M}(i,j;h,\ell)$ via Procedure (b). The corresponding three cases are labeled by from left to right by J_1, J_2 and J_3, respectively.

$$Q_{\triangle,\triangledown,\square}^{\mathsf{K,M}}(i,i_1;r,j_1) = Q_{\triangle}^{\mathsf{K,M}}(i,i_1;r,j_1) + Q_{\triangledown}^{\mathsf{K,M}}(i,i_1;r,j_1) + Q_{\square}^{\mathsf{K,M}}(i,i_1;r,j_1),$$

the recursion of the partition function for $J_{DT}^{\mathsf{K,M}}(i,j;h,\ell)$ is given by:

$$(4.2) \qquad\qquad Q_{DT}^{\mathsf{K,M}}(i,j;r,s) = \sum_{i_1,j_1}(Q(J_1) + Q(J_2) + Q(J_3)),$$

where

$$
\begin{aligned}
Q(J_1) &= Q^{\mathsf{K,M}}_{\triangle,\triangledown,\square}(i,i_1;r,j_1)Q^{\mathsf{K,M}}_{DT}(i_1+1;j_1+1,s) \\
Q(J_2) &= Q^{\mathsf{F,M}}_{\triangle,\triangledown,\square}(i,i_1;r,j_1)Q^{\mathsf{K,M}}_{DT}(i_1+1;j_1+1,s)) \\
Q(J_3) &= Q^{\mathsf{K,M}}_{\triangle,\triangledown,\square}(i,i_1;r,j_1)Q^{\mathsf{F,M}}_{DT}(i_1+1;j_1+1,s).
\end{aligned}
$$

Case 3: $J^{\mathsf{K,K},B,B}_{DT}(i,j;h,\ell)$. By definition of $J^{\mathsf{K,K},B,B}_{DT}(i,j;h,\ell)$, we have $(R[j],S[\ell]) \in J(i,j;h,\ell)$. We consider the set of exterior arcs in $J(i,j-1;r,\ell-1)$, W. In case of $W = \varnothing$, $J^{\mathsf{K,K},B,B}_{DT}(i,j;h,\ell)$ decomposes into $R[i,j-1]$, $S[h,\ell-1]$ and $(R[j],S[\ell])$. This is the leftmost (first) case (L_1) displayed in Fig. 25. Otherwise, let $(R[i_1],S[j_1])$ denote the maximal exterior arc in $J(i,j-1;r,\ell-1)$. We consider the unique $J(i,j;r,\ell)$-tight structure which contains $(R[i_1],S[j_1])$, denoted by $J_T(R[i_1],S[j_1])$. If $J_T(R[i_1],S[j_1])$ has not type $\circ$, we have the second case (L_2) displayed in Fig. 25. Otherwise, depending on whether or not $R[j_1+1,j-1] = \varnothing^{j-1}_{j_1+1}$ and $S[h_1+1,\ell-1] = \varnothing^{\ell-1}_{h_1+1}$, we have the third (L_3) and fourth case (L_4), displayed in Fig. 25. Consequently, we arrive at:
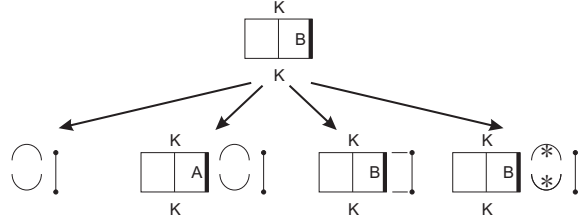

FIGURE 25. The four decomposition scenarios of $J^{\mathsf{K,K},B,B}_{DT}(i,j;r,s)$ via Procedure (b). We denote the corresponding cases from left to right by L_1, L_2 L_3 and L_4, respectively.

$$
\text{(4.3)} \qquad Q^{\mathsf{K,K},B,B}_{DT}(i,j;h,\ell) = \sum_{j_1,h_1}(Q(L_1)+Q(L_2)+Q(L_3)+Q(L_4)),
$$

where $Q(L_1) = Q^{\mathsf{k}}_s(i,j-1)Q^{\mathsf{k}}_s(h,\ell-1)$ and

$$
\begin{aligned}
Q(L_2) &= Q^{\mathsf{K,K},B,A}_{RT}(i,j_1;h,h_1)(Q^{\mathsf{k}}(j_1+1,j-1)+Q^{\mathsf{k}}(h_1+1,\ell-1)+Q^{\mathsf{k}}(j_1+1,j-1)Q^{\mathsf{k}}(h_1+1,\ell-1)) \\
Q(L_3) &= Q^{\mathsf{K,K},B,B}_{RT}(i,j_1;h,h_1)e^{-\sigma_0+\sigma\mathsf{Int}(j,\ell;j_1,h_1)} \\
Q(L_4) &= Q^{\mathsf{K,K},B,B}_{RT}(i,j_1;h,h_1)(Q^{\mathsf{k}}(j_1+1,j-1)+Q^{\mathsf{k}}(h_1+1,\ell-1)+Q^{\mathsf{k}}(j_1+1,j-1)Q^{\mathsf{k}}(h_1+1,\ell-1))
\end{aligned}
$$

5. BASE PAIRING PROBABILITIES

We have seen in Section 3 that the probability of a joint structure, $J(1, N; 1, M)$, is given by

$$(5.1) \qquad \mathbb{P}(J(1, N; 1, M)) = \frac{1}{Q^I} e^{-F(J(1,N;1,M))/kT},$$

where $Q^I = \sum_{J(1,N;1,M)} e^{-F(J(1,N;1,M))/kT}$. In this section, we shall calculate the base pair probabilities (BPP) for interior and exterior arcs. The key idea is here to associate the probability of specific substructures contained in the decomposition tree. In other words, a term $Q_\xi^{Y_1,Y_2,Y_3,Y_4}(i, j; h, \ell)$ in the recursive calculation of the partition function gives rise to the probability $\mathbb{P}_\xi^{Y_1,Y_2,Y_3,Y_4}(i, j; h, \ell)$. For instance, $\mathbb{P}_{RT}^{M,K,A,B}(i, j; h, \ell)$ is, by construction, the sum over all the probabilities of joint structures $J(1, N; 1, M)$ such that $J(i, j; h, \ell)$ is contained in $T(J(1, N; 1, M))$ and $J(i, j; h, \ell) = J_{RT}^{M,K,A,B}(i, j; h, \ell)$. We remark that the above observations reduce the computation of the BPP to a trace-back routine in the decomposition tree, constructed in Section 3.

The basic strategy can be sketched as follows:
(a) derive from the recursion of the partition function the corresponding recursion of the probabilities
(b) partition the substructures according to their respective contribution to the partition function
(c) for each subclass, recursively calculate the probability of substructures via tracing back the decomposition tree.

We recall that $\Sigma_0 = \{J(1, N; 1, M) \mid J(i, j; h, \ell) \in T(J(1, N; 1, M))\}$. The probability $\mathbb{P}(i, j; h, \ell)$ is given by

$$(5.2) \qquad \mathbb{P}(i, j; h, \ell) = \sum_{J(1,N;1,M)\in\Sigma_0} \mathbb{P}(J(1, N; 1, M)).$$

We accordingly set

$$(5.3) \qquad \mathbb{P}_\xi^{Y_1,Y_2,Y_3,Y_4}(i, j; h, \ell) = \sum_{J(1,N;1,M)\in\Lambda_0} \mathbb{P}(J(1, N; 1, M)),$$

where $\Lambda_0 = \{J(1, N; 1, M) \mid J(i, j; h, \ell) \in T(J(1, N; 1, M)), J(i, j; h, \ell) \in J_\xi^{Y_1,Y_2,Y_3,Y_4}(i, j; h, \ell)\}$.

5.1. Base pairing probabilities for RNA secondary structures.
In order to illustrate the concept, let us put the calculation of the BPP for secondary structures into the context of our

backtracking routine. Given a secondary structure R of length N, the probability of R is given by $\mathbb{P}(R) = \frac{1}{Q}e^{-F(R)/kT}$. In order to calculate the probability of $R[i]$ being connected to $R[j]$ in the equilibrium ensemble of structures, $\mathbb{P}(i_R, j_R)$, the first objective is to express the probability of this base pair into a sum of probabilities of substructures. Let $T(R[1, N])$ be the decomposition tree of a particular secondary structure $R[1, N]$ via Procedure (c) and $\Omega(i_R, j_R) = \{S \mid (R[i], R[j]) \in T(S)\}$. We remark that $\Omega(i_R, j_R)$ coincides set of secondary structure such that $R[i]$ is bound with $R[j]$, see Section 3, Observation 2. Then we have

$$(5.4) \qquad \mathbb{P}(i_R, j_R) = \frac{\sum_{S \in \Omega(i_R, j_R)} Q(S)}{Q}.$$

Let $R^b(i, j)$ denote the set of segments $R[i, j]$ in which $R[i]$ is connected with $R[j]$ and $R[i, j] \in T(R[1, N])$. By construction, $\mathbb{P}^b(i_R, j_R)$ is the probability of $R^b(i, j)$. According to Procedure (c), we have $\mathbb{P}(i_R, j_R) = \mathbb{P}^b(i_R, j_R)$ since $(R[i], R[j]) \in T(J(1, N; 1, M))$ if and only if the parent of $(R[i], R[j])$ in the decomposition tree belongs to $R^b(i, j)$. Therefore the problem is reduced to the calculation of $\mathbb{P}^b(i_R, j_R)$. Inspection of Procedure (c) shows, that for the parent of an element of $R^b(i, j)$ we have to distinguish the five cases displayed in Fig. 26. Let $R^m(i, j)$ denote the set
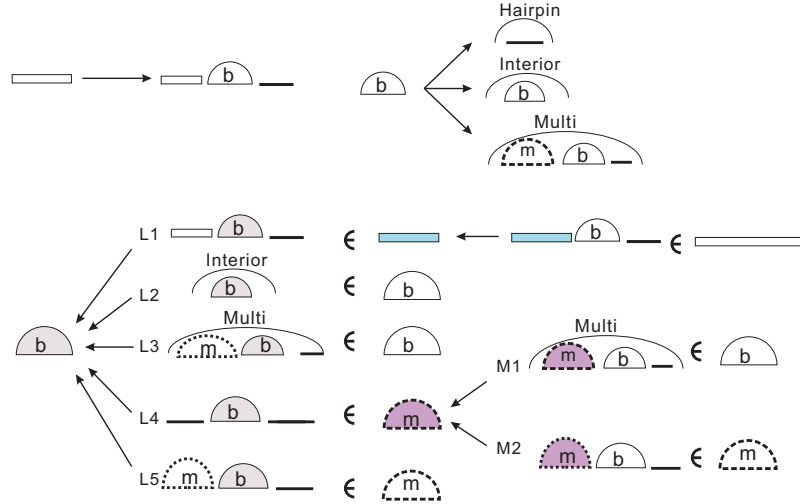

FIGURE 26. Tracing back: for a parent of $R^b(i, j)$ we have according to Procedure (c), five cases, labeled from top to bottom by L_1, L_2, L_3, L_4 and L_5. For a parent of $R^m(i, j)$ there are two cases, denoted by M_1 and M_2.

of segments $R[i, j] \in T(R[1, N])$ such that $R[i, j] \neq \varnothing_i^j$, where the outer loop has type M. Let $R^s(i, j)$ denote the set of segments $R[i, j] \in T(R[1, N])$. In particular, $R^s(1, N) = R[1, N]$. Set

$\mathbb{P}^m(i_R, j_R)$ and $\mathbb{P}^s(i_R, j_R)$ be the probability of $R^m(i,j)$ and $R^s(i,j)$, respectively. Then we have $\mathbb{P}^b(i,j) = \mathbb{P}(L_1) + \mathbb{P}(L_2) + \mathbb{P}(L_3) + \mathbb{P}(L_4) + \mathbb{P}(L_5)$, where

$$
\begin{aligned}
\mathbb{P}(L_1) &= \sum_{h,\ell} \mathbb{P}^s(h,\ell) \frac{Q^s(h, i-1) Q^b(i,j)}{Q^s(h,\ell)} \\
\mathbb{P}(L_2) &= \sum_{h,\ell} \mathbb{P}^b(h,\ell) \frac{Q^b(i,j) e^{-\mathsf{Int}(i,j;h,\ell)/kT}}{Q^b(h,\ell)} \\
\mathbb{P}(L_3) &= \sum_{h,\ell} \mathbb{P}^b(h,\ell) \frac{Q^b(i,j) e^{-(\alpha_1 + 2\alpha_2 + (\ell-j-1)\alpha_3)/kT} Q^m(h+1, i-1)}{Q^b(h,\ell)} \\
\mathbb{P}(L_4) &= \sum_{h,\ell} \mathbb{P}^m(h,\ell) \frac{Q^b(i,j) e^{-((i-h+\ell-j)\alpha_3 + \alpha_2)/kT}}{Q^m(h,\ell)} \\
\mathbb{P}(L_5) &= \sum_{h,\ell} \mathbb{P}^m(h,\ell) \frac{Q^b(i,j) e^{-((\ell-j)\alpha_3 + \alpha_2)/kT} Q^m(h, i-1)}{Q^m(h,\ell)} .
\end{aligned}
$$

Accordingly, the recurrence formulae for $\mathbb{P}^m(i,j)$ and $\mathbb{P}^s(i,j)$ are given as follows:

$$
\begin{aligned}
\mathbb{P}^m(i,j) &= \mathbb{P}(M_1) + \mathbb{P}(M_2) \\
\mathbb{P}(M_1) &= \sum_{h,\ell} \mathbb{P}^b(i-1,\ell) \frac{e^{-(\alpha_1 + 2\alpha_2 + (\ell-1-h)\alpha_3)/kT} Q^b(j+1, h) Q^m(i,j)}{Q^b(i-1,\ell)} \\
\mathbb{P}(M_2) &= \sum_{h,\ell} \mathbb{P}^m(i,\ell) \frac{Q^m(i,j) Q^b(j+1, h)}{Q^m(i,\ell)} \\
\mathbb{P}^s(i,j) &= \sum_{h,\ell} \mathbb{P}^s(i,\ell) \frac{Q^s(i,j) Q^b(j+1, h)}{Q^s(i,\ell)} .
\end{aligned}
$$

5.2. **Base pairing probabilities for joint structures.** Following the basic strategy, we first express the BPP via the probabilities of particular substructures. In the following, we abbreviate $J(1, N; 1, M)$ by J. In order to calculate $\mathbb{P}(i_R, j_R)$, let $\Sigma_1 = \{J \mid (R[i], R[j]) \in J\}$, we consider the parent of $(R[i], R[j])$ in the $T(J)$ and accordingly obtain

$$(5.5) \quad \Sigma_1 = \{J \mid R^b[i,j] \in T(J)\} \cup \bigcup_{h,\ell} \{J \mid J_{\triangledown}(i,j;h,\ell) \in T(J)\} \cup \bigcup_{h,\ell} \{J \mid J_{\square}(i,j;h,\ell) \in T(J)\},$$

which immediately leads to

$$(5.6) \qquad \mathbb{P}(i_R, j_R) = \mathbb{P}^b(i_R, j_R) + \sum_{h,\ell} \mathbb{P}_{\triangledown}(i,j;h,\ell) + \sum_{h,\ell} \mathbb{P}_{\square}(i,j;h,\ell),$$

where

$$(5.7) \qquad \mathbb{P}_{\triangledown}(i,j;h,\ell) = \mathbb{P}^E_{\triangledown}(i,j;h,\ell) + \mathbb{P}^M_{\triangledown}(i,j;h,\ell) + \mathbb{P}^K_{\triangledown}(i,j;h,\ell) + \mathbb{P}^F_{\triangledown}(i,j;h,\ell),$$

$$(5.8) \qquad \mathbb{P}_{\square}(i,j;h,\ell) = \mathbb{P}^E_{\square}(i,j;h,\ell) + \mathbb{P}^M_{\square}(i,j;h,\ell) + \mathbb{P}^K_{\square}(i,j;h,\ell) + \mathbb{P}^F_{\square}(i,j;h,\ell).$$

Analogously, for $\mathbb{P}(i_S, j_S)$ we set

$$(5.9) \qquad \Sigma_2 = \{J \mid S^b[h,\ell] \in T(J)\} \cup \bigcup_{i,j}\{J \mid J_{\triangle}(i,j;h,\ell) \in T(J)\},$$

and obtain

$$(5.10) \qquad \mathbb{P}(i_S, j_S) = \mathbb{P}^b(h_S, \ell_S) + \sum_{h,\ell} \mathbb{P}_{\triangle}(h,\ell;i,j),$$

where

$$(5.11) \qquad \mathbb{P}_{\triangle}(h,\ell;i,j) = \mathbb{P}^E_{\triangle}(h,\ell;i,j) + \mathbb{P}^M_{\triangle}(h,\ell;i,j) + \mathbb{P}^K_{\triangle}(h,\ell;i,j) + \mathbb{P}^F_{\triangle}(h,\ell;i,j).$$

We remark that the expressions for the BPP $\mathbb{P}(i_R, j_R)$ and $\mathbb{P}(i_S, j_S)$ are not symmetric. This is due to the fact that in our decomposition routines always the outer arcs contained in R are given preference. In other words, the asymmetry is a result of our particular construction. Finally, we calculate the binding probability of an exterior arc $(R[i], S[j])$. Since $(R[i], S[j])$, being a tight structure of type $\circ$, is already substructure, we can skip the first two steps of the basic strategy. In order to compute the binding probabilities of both: interior and exterior arcs, the key is to employ an "inverse" grammar induced by tracing back in the decomposition tree as displayed in Fig. 27. By virtue of this backtracking, we obtain the recurrence formulae in analogy to the case of secondary structures, discussed above.

6. Synopsis

In this paper we derive the partition function and the base pairing probabilities of RNA interaction structures. Furthermore we present the algorithm rip that computes the partition function and the base pairing probabilities in $O(N^4 M^2) + O(N^2 M^4)$ time and $O(N^2 M^2)$ space.

While the partition function is due to [11] our construction is independently derived and based on two ideas: the concept of tight structure in Section 1 and the decomposition tree, presented in Section 3. We did however, adopt the notions of kissing and hybrid loops from [11]. The derivation of the base pairing probabilities for joint structures is new. Here the key idea is to express the

latter via energy-wise "quantifiable" substructures, that are contained in the decomposition tree. We discussed that in contrast to the computation of the base pairing probabilities of secondary structures, the specific construction of the unique grammar factors in. As a result, being a joint substructure containing a certain base pair, is not the correct criterion any more. Only those substructures that are obtained via tracing back in the decomposition tree contribute to the base pairing probability.

The complete set of partition function recursions and all details on the particular implementation of rip can be found at

<p align="center">http://www.combinatorics.cn/cbpc/rip.html</p>

Finally, we also compute the generating function of joint structures. The analysis of this function is beyond the scope of this paper and can be found as supplemental material at the above web-site.

References

[1] T. Akutsu. Dynamic programming algorithms for RNA secondary structure prediction with pseudoknots. *Disc. Appl. Math.*, 104:45–62, 2000.

[2] C. Alkan, E. Karakoc, J.H. Nadeau, S.C. Sahinalp, and K.Z. Zhang. RNA-RNA interaction prediction and antisense RNA target search. *J. Comput. Biol.*, 13:267–282, 2006.

[3] M. Andronescu, Z.C. Zhang, and A. Condon. Secondary structure prediction of interacting RNA molecules. *J. Mol. Biol.*, 345:1101–1112, 2005.

[4] J.P. Bachellerie, J. Cavaillé, and A. Hüttenhofer. The expanding snoRNA world. *Biochimie*, 84:775–790, 2002.

[5] D. Banerjee and F. Slack. Control of developmental timing by small temporal RNAs: a paradigm for RNA-mediated regulation of gene expression. *Bioessays*, 24:119–129, 2002.

[6] R. Benne. RNA editing in trypanosomes. the use of guide RNAs. *Mol. Biol.Rep.*, 16:217–227, 1992.

[7] S. Bernhart, H. Tafer, U. Mückstein, C. Flamm, P.F. Stadler, and I.L. Hofacker. Partition function and base pairing probabilities of RNA heterodimers. *Algorithms Mol. Biol.*, 1:3–3, 2006.

[8] A. Busch, A.S. Richter, and R. Backofen. IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics*, 24:2849–2856, 2008.

[9] W.Y.C. Chen, J. Qin, and C.M. Reidys. Crossings and nestings in tangled diagrams. *Electron. J. Comb.*, 15:R86, 2008.

[10] W.Y.C. Chen, J. Qin, C.M. Reidys, and D. Zeilberger. Efficient counting and asymptotics of k-noncrossing tangled diagrams. *Electron. J. Comb.*, 16:R37, 2009.

[11] H. Chitsaz, R. Salari, S.C. Sahinalp, and R. Backofen. A partition function algorithm for interacting nucleic acid strands. 2009.

[12] A. Fire, S. Xu, and Kostas S.A. Montgomery, M.K. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature*, 391:806–811, 1998.

[13] S.M. Hammond, E. Bernstein, D. Beach, and G.J. Hannon. An RNA-directed nuclease mediates pose-transcriptional gene sliencing in drosophlia cells. *Nature*, 404:293–296, 2000.

[14] I.L. Hofacker, W. Fontana, P.F. Stadler, L.S. Bonhoeffer, M. Tacker, and P. Schuster. Fast folding and comparison of RNA secondary structures. *Monatsh. Chem.*, 125:167–188, 1994.

[15] J A Howell, T F Smith, and M S Waterman. Computation of generating functions for biological molecules. *J Appl Math*, 39:119–133, 1980.

[16] E.Y. Jin, J. Qin, and C.M. Reidys. Combinatorics of RNA structures with pseudoknots. *J. Math. Biol.*, 70:45–67, 2008.

[17] E.Y. Jin and C.M. Reidys. Combinatorial design of pseudoknot RNA. *Adv. Appl. Math.*, 42:135–151, 2009.

[18] J. Kugel and J. Goodrich. An RNA transcriptional regulator templates its own regulatory RNA. *Nat. Struct. Mol. Biol.*, 3:89–90, 2007.

[19] D. Mathews, J. Sabina, M. Zuker, and D.H. Turner. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.*, 288:911–940, 1999.

[20] J.S. McCaskill. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, 29:1105–1119, 1990.

[21] M.T. McManus and P.A. Sharp. Gene silencing in mammals by small interfering RNAs. *Nature Reiviews*, 3:737–747, 2002.

[22] U. Mückstein, H. Tafer, S.H. Bernhard, M. Hernandez-Rosales, J. Vogel, P.F. Stadler, and I.L. Hofacker. Translational control by RNA-RNA interaction: Improved computation of RNA-RNA binding thermodynamics. In Mourad Elloumi, Josef Küng, Michal Linial, Robert F. Murphy, Kristan Schneider, and Cristian Toma Toma, editors, *BioInformatics Research and Development — BIRD 2008*, volume 13 of *Comm. Comp. Inf. Sci.*, pages 114–127, Berlin, 2008. Springer.

[23] U. Mückstein, H. Tafer, J. Hackermüller, S.H. Bernhard, P.F. Stadler, and I.L. Hofacker. Thermodynamics of RNA-RNA binding. *Bioinformatics*, 22:1177–1182, 2006. Earlier version in: *German Conference on Bioinformatics 2005*, Torda, Andrew and Kurtz, Stefan and Rarey, Matthias (eds.), Lecture Notes in Informatics **P-71**, pp 3-13, Gesellschaft f. Informatik, Bonn 2005.

[24] A. Nykanen, B. Haley, and P.D. Zamore. ATP requirements and small interfering RNA structure in the RNA interference. *Cell*, 107:309–321, 2001.

[25] S. Parrish, J. Fleenor, S. Xu, C. Mello, and A. Fire. A functional anatomy of a dsRNA trigger: Differential requirement for the two trigger strands in RNA interference. *Mol. Cell.*, 6:1077–1087, 2000.

[26] J. Qin and C.M. Reidys. A framework for RNA tertiary interaction. 2008.

[27] B.J. Reinhart, F.J. Slack, M. Basson, A.E. Pasquinelli, J.C. Bettinger, A.E. Rougvie, H.R. Horvitz, and G. Ruvkun. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabdites elegans. *Nature*, 403:901–906, 2000.

[28] J. Ren, B. Rastegari, A. Condon, and H.H. Hoos. Hotknots: heuristic prediction of microRNA secondary structures including pseudoknots. *RNA*, 11:1494–1504, 2005.

[29] E. Rivas and S.R. Eddy. A dynamic programming algorithms for RNA structure prediction including pseudoknots. *J. Mol. Biol.*, 285:2053–2068, 1999.

[30] W.R. Schmitt and M.S. Waterman. Linear trees and RNA secondary structure. *Disc. Appl. Math.*, 51:317–323, 1994.

[31] E.G.H. Wagner and K. Flardh. Antisense RNAs everywhere? *Thends Genet.*, 244:48–52, 2002.

[32] M.S. Waterman and T.F. Smith. RNA secondary structure: A complete mathematical analysis. *Math. Biosci*, 42:257–266, 1978.

[33] M.S. Waterman and T.F. Smith. Rapid dynamic programming algorithms for RNA secondary structure. *Adv. Appl. Math.*, 7:455–464, 1986.

[34] D. Yang, H. Lu, and J.W. Erickson. Evidence that processed small dsRNA may mediate sequence-specific mRNA degradation during RNAi in drosophila embryos. *Curr. Biol.*, 10:1191–1200, 2000.

[35] D.A. Zorio, K. Lea, and T. Blumenthal. Cloning of caenorhabditis u2af65: an alternatively spliced RNA containing a novel exon. *Mol. Cell. Biol.*, 17:946–953, 1997.

[1]CENTER FOR COMBINATORICS, LPMC-TJKLC, NANKAI UNIVERSITY, TIANJIN 300071, P.R. CHINA, PHONE: *86-22-2350-6800, FAX: *86-22-2350-9272, [2]UNIVERSITY OF LEIPZIG, LEIPZIG D-04107, HÄRTELSTR. 16-18, GERMANY,
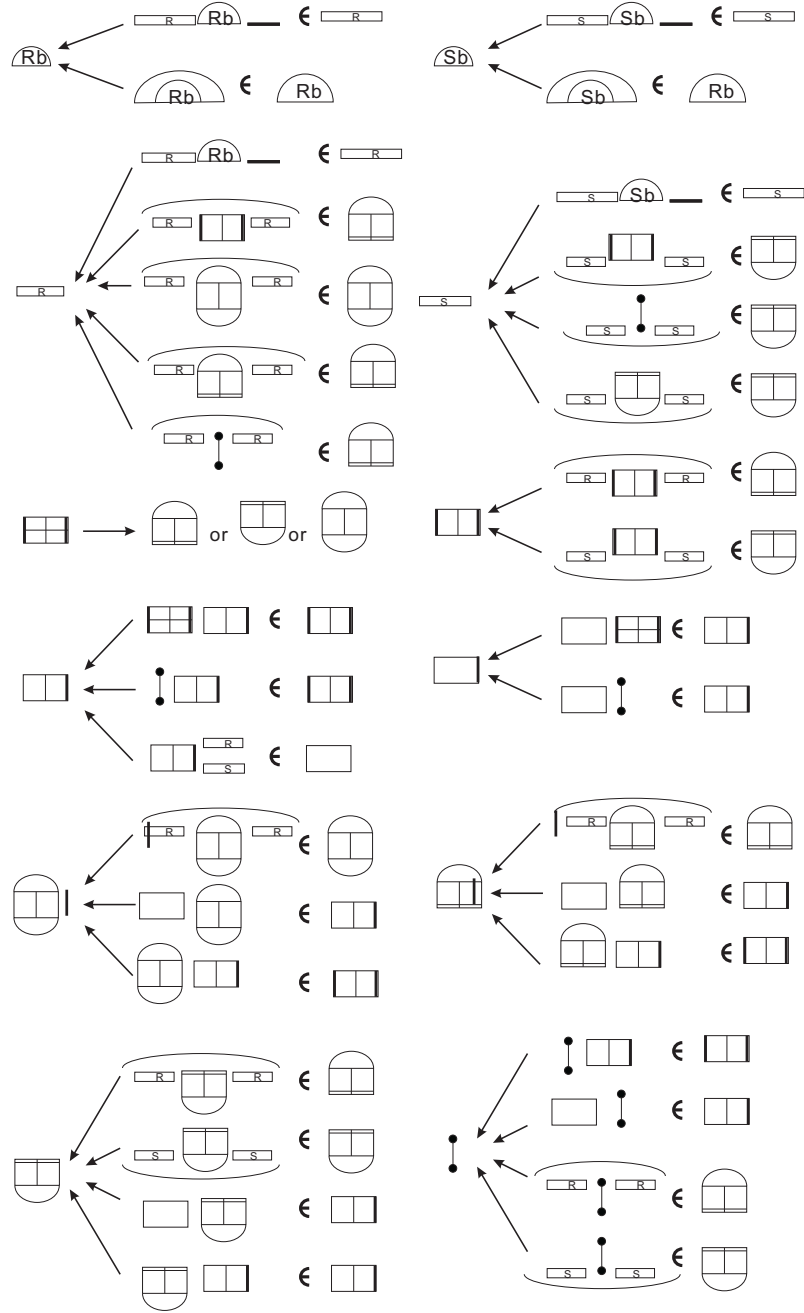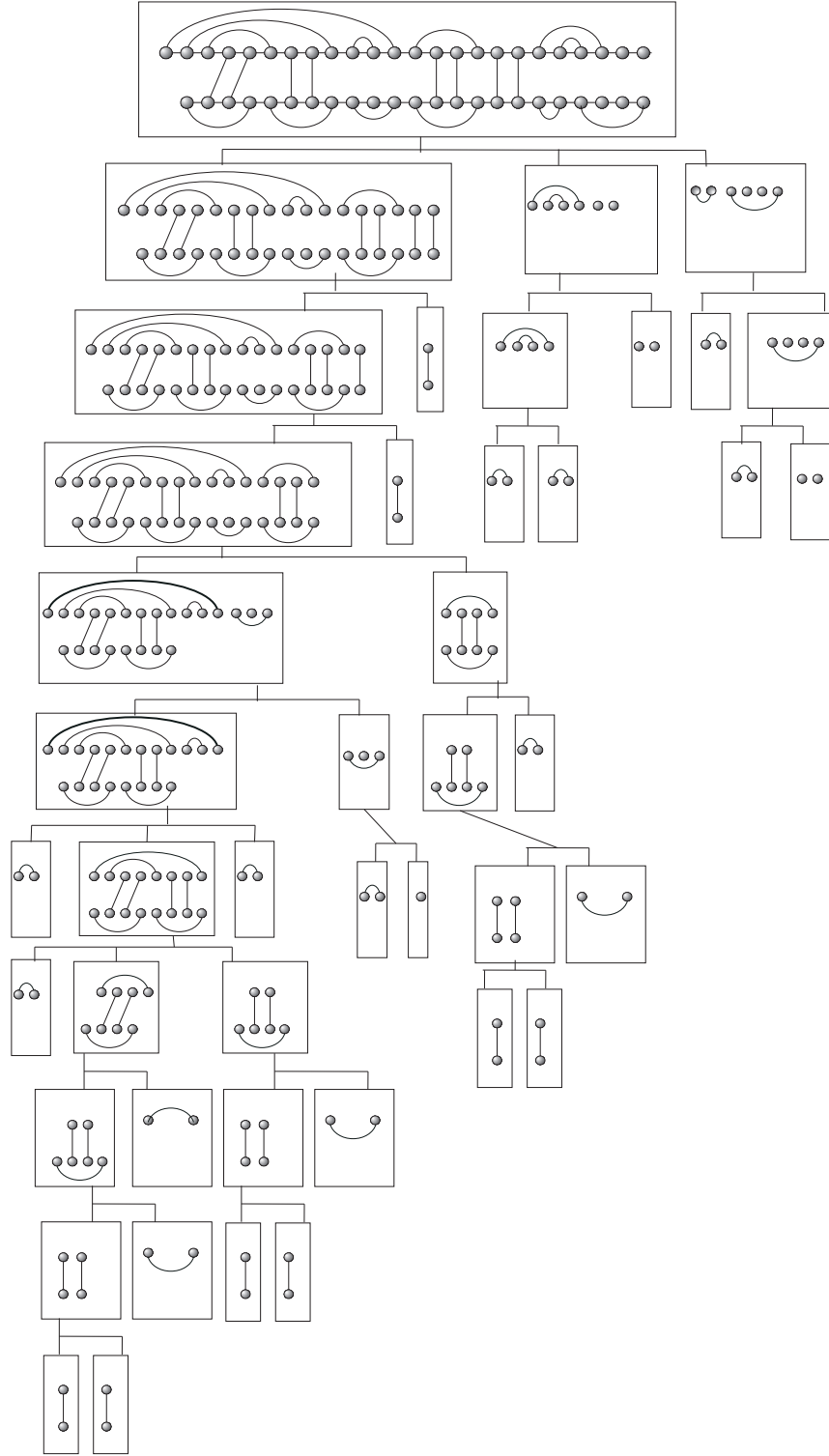
E-mail address: duck@santafe.edu

FIGURE 27. Illustration of the "inverse" grammar, obtained by back-tracing in the decomposition tree.

FIGURE 28. The decomposition tree $T(1, N; 1, M)$.